Randomness: a property of the mathematical and physical systems

Anastasios A. Tsonis

To cite this article: Anastasios A. Tsonis (2016) Randomness: a property of the mathematical and physical systems, Hydrological Sciences Journal, 61:9, 1591-1610, DOI: 10.1080/02626667.2014.992434

To link to this article: https://doi.org/10.1080/02626667.2014.992434

Accepted author version posted online: 27 Nov 2014.
Published online: 07 Apr 2016.

Submit your article to this journal

Article views: 161

View Crossmark data

Citing articles: 1 View citing articles
Randomness: a property of the mathematical and physical systems

Anastasios A. Tsonis

Department of Mathematical Sciences, Atmospheric Sciences Group, University of Wisconsin-Milwaukee, Milwaukee, WI 53201, USA

ABSTRACT

This article is a concept paper, which discusses the definition of randomness, and the sources of randomness in the mathematical system as well as in the physical system (the Universe). We document that randomness is an inherited property of mathematics and of the physical world, shaping all observed forms and structures, and we discuss its role.

ARTICLE HISTORY

Received 12 June 2014
Accepted 16 October 2014

EDITOR

D. Koutsoyiannis

GUEST EDITOR

E. Volpi

KEYWORDS

randomness; fractals; chaos; irreversibility; stochasticity

Introduction

"Others there are who believe that chance is a cause but that it is inscrutable to human intelligence as being a divine thing and full of mystery."

Aristotle, Physics Book II, 4

Over 2500 years ago Aristotle pondered on what many other philosophers have pondered throughout time. Exactly what is randomness and why is it there? This paper presents a synthesis of facts from the mathematical and physical systems, which clearly establish randomness as a property of nature and that what we see and experience around us emerges from the interplay of rules and randomness. This paper is arranged in four parts. Part 1 deals with randomness in the mathematical system. Part 2 deals with randomness in the physical system (the Universe). In Part 3, the connections between the sources of randomness in these two systems is discussed. Finally, in Part 4, we discuss the role of randomness in the physical world. For more detailed discussions the reader is referred to (Tsonis 2008).

Prologue

Consider an equilateral triangle of side size $L$ and the following iteration: take the middle third of each side and replace it with two $L/3$ length sides forming a smaller equilateral triangle on each original side. We now have a ‘star’ with 12 sides. Repeat the process for each of the 12 sides, and keep on repeating for the new sides ad infinitum. This process will result in a closed boundary, which is called the Koch island or Koch snowflake (Fig. 1(a)). This boundary is an exact fractal (Mandelbrot 1983), but a far cry from real boundaries such as coastlines. We can ‘improve’ on that boundary by introducing randomization of the iteration process; for example rather than forming the equilateral triangle with the same orientation at each step, we may choose the orientation at random. This leads to a boundary that is an improvement in the right direction Fig. 1(b), but still it is a far cry from the actual coastline. However, we only need to be a little more creative with our randomization technique (Peitgen et al. 1988, 1992) before we can generate a random fractal whose details will be indistinguishable from natural coastlines. One such example is shown in Fig. 1(c). Comparison with the coast of England shows striking similarities at all scales (Mandelbrot 1983).

Let us now consider an example of hydrological importance. Figure 2 is taken from Rodriguez-Iturbe and Rinaldo (1997) and refers to fractal river networks. Figure 2(a) shows the river basin from a digital elevation model (DEM) of the Sarca di Nambrone basin in Italy. The area of the basin is 21.5 km$^2$ and it is delineated on a grid of 2151 pixels of size $100 \times 100$ m$^2$. River basins are composed of the channel network and the hillslopes. The hillslopes control the runoff and the network transports the runoff to some outlet. River basins develop a tree-like structure where every branch in the network
is connected to a downstream branch. This leads to an efficient and effective drainage system. Figure 2(b) is a corresponding optimal channel network (OCN). This computer-generated river basin is grown from a random initial configuration within the same boundaries as in Fig. 2(a). The growth process involves iteration processes and randomization (Rodriguez-Iturbe and Rinaldo 1997). The physical basis of OCNs is the minimization of energy expenditure, which calls for the minimization of a functional relationship involving the rate of optimal energy expenditure of each branch, its mean annual flow, and its length. Clearly, the real and the computer-generated basin are very similar. In fact, as it is stated in Rodriguez-Iturbe and Rinaldo (1997) they are both random fractals with the same fractal dimension. The key here is that this structure is the result of a rule (the minimization of energy expenditure) and randomness. Indeed, this theme is of great importance in the hydrological community (Koutsoyiannis et al. 2009, Koutsoyiannis 2010), as well as in many other areas of science (Tsonis 2008).
Exact fractals, like Euclidean structures (straight lines, spheres, cubes, etc.) are almost never observed in nature. Randomness eliminates such possibilities. This is also supported by the study of cellular automata. Cellular automata (Wolfram 2002) are systems whose evolution is described not by equations but by very simple, computer-program-like rules. They can provide an alternative to more complicated systems described by differential equations. By studying thousands of completely deterministic cellular automata, Wolfram identified five types of evolution: steady states, periodic structures, exact fractals, chaotic evolutions, and evolutions characterized by a mixture of regular and irregular structures (often referred to as the edge of chaos). When, however, randomness is introduced to the automata, the exact fractals class does not emerge (see also Tsonis 1996). Clearly then, unless randomness is combined with rules, no realistic forms of natural phenomena will emerge. Numerous other examples can be given from all areas of science to support this statement. Then the obvious question arises: what is randomness and where is it coming from?

Part 1: Randomness in the mathematical system

In a seminal paper in 1931, Gödel proved that there are mathematical statements that cannot be proved within the current mathematical system. Gödel proved that if all mathematical statements could be proved (which will indicate that the formal mathematical system is complete) then the system will be an inconsistent system. This self-reference about the mathematical system also proves that consistent mathematics is an incomplete system. This means that in a consistent mathematical system there will always be uncertainty about certain statements. This uncertainty introduces a form of randomness into the formal mathematical system. Formally, Gödel’s Incompleteness Theorem, is expressed as:

For every consistent formalization of arithmetic, no matter how complex, there exist arithmetic truths improvable within that formal system,

or in a somewhat simpler form:

In today’s mathematics there are true statements about numbers that cannot be proved.

A nice everyday example of the principle underlying Gödel’s theorem is given by Douglas Hofstadter in his monumental book Gödel, Escher, Bach: an Eternal Golden Braid (Hofstadter 1979). Consider a phonograph, which is playing a record in a room. The phonograph produces sounds, which are sent out to its surrounding environment. A sound is a vibration. These vibrations, as well as other vibrations from other sources, are reflected by the walls and propagate back to the phonograph. In this way the reflected vibrations may affect the phonograph’s operation. Obviously, the stronger the vibrations a record produces the greater the effect they have on the phonograph. As such, for any record player there may be records which cannot be played because they may cause its indirect self-destruction.

The system of mathematics can become less incomplete by adding more rules. There are numerous cases where mathematical statements were proven only after new insights (new rules) were discovered. The Fermat Conjecture is the most celebrated of such examples. Proposed by Pierre Fermat in 1665, it states that the equation $x^n + y^n = z^n$ (a Diophantine equation) does not have a positive integer solution when $n$ is an integer >2. Thousands of mathematicians wrestled with this problem unsuccessfully until Andrew Wiles finally proved it in 1995. Why did it take 330 years before anyone could prove the conjecture? Because there were areas of mathematics, specifically the theory of elliptic curves, which had to be discovered before anyone could prove the conjecture.

However, unless an infinite number of rules are added one cannot be certain that the system will not be incomplete. What that means is that there is no finite set of rules that can be consistently added to the system to make it complete. A consequence of this is that a procedure, which decides that any mathematical statement is true in a finite number of steps, does not exist. Think of the system as a photograph. Many years ago photographs were black and white and blurry. As such there were ‘truths’ (for example, the color of the sky) that could not be ‘proven’ by them. As technology improved (read: more rules were added), photographs became more realistic (or more complete). Still, however, unless we have an infinite resolution, the details in a scene that is being photographed cannot be known exactly (for example, individual molecules cannot be seen).

Since the time of Euclid, the dream of mathematicians was to reduce mathematics to a set of basic axioms from which, through inference, all theorems could be proven. Gödel’s theorem shook the foundations of mathematics by showing that this is not possible. The implications are startling. Apart from philosophical issues such as “can we ever know the truth from reasoning?”, Gödel’s theorem implies that whether a mathematical statement is true or false may not be known. Moreover, since there can be an infinity
of such statements, Gödel’s theorem implies that the element of uncertainty, and thus randomness, is interwoven with axioms, theorems, and the whole structure of mathematics. This naturally brings up the following question: what exactly is randomness?

The earliest mathematical treatment of randomness was attempted by Blaise Pascal, Pierre de Fermat, and Christian Huygens in the 17th century. Fundamental to their approach is that outcomes of random processes are equally probable. This is considered the first definition of randomness in statistical terms. While this definition may be appropriate to processes like tossing a coin, it is not adequate for many other processes. Think, for example, of earthquakes as random events. Very powerful earthquakes are not equally probable with weak events. By the early 1900s, mathematicians were beginning to wrestle with new definitions of randomness; among the first was Richard Von Mises. According to Von Mises, randomness is the inability to devise a system that will allow us to predict where in the sequence a particular observation will occur without prior knowledge of the whole sequence. This amounts to saying that a sequence is random if there is no rule or law that can be used to produce it (or that each step is completely independent of the previous step). Von Mises definition, however, runs into problems when it comes to irrational numbers. The fact is that irrational numbers such as \( \pi \), \( \sqrt{2} \), and \( e \), which, even though they are aperiodic and for all practical purposes random (without an obvious pattern), can be generated by simple algorithms and as such the value of a digit at a given place is known. This fact suggests that all mathematical objects obey some law, which simply means that there exists no sequence that conforms to no law. It follows that random sequences à la Von Mises may not exist or that perhaps the notion of randomness itself is inconsistent.

Von Mises’ definition implies that every rule or law will fail to predict an observation in a random sequence. In the early 1960s, Andrei Kolmogorov (in Russia), Gregory Chaitin (in the USA), and Ray Solomonoff (in the USA) were independently wrestling with a new view of randomness. Their attempts to improve Von Mises’ definition begin with the idea of relaxing the condition “every rule or law”. We will introduce their definition of randomness with a series of simple examples. First let us consider the binary sequence:

\[ 010101010101010101010101 \]

We only need to look at this sequence to realize (without any mathematical analysis) that it is periodic. In fact we can reproduce this sequence by writing a rule or a computer ‘program’ that, for illustration purposes, is one line long or, if you prefer, requires 21 keystrokes to type it (including the spaces between words):

Print 01 twelve times

It follows that all periodic sequences can be described by very simple rules or very short programs. Next let us consider the following sequence, which is 25 digits long:

\[ 1101100011010011010100110 \]

In this case it would take more than a second to find the rule that produces this sequence. For all practical purposes this sequence appears random. There is no apparent pattern, which will enable us to write a simple program in this case. Thus, the program that produces the sequence might as well have as many lines as the number of digits:

Print 1
Print 1
Print 0
...
...
...
Print 0

This program is ‘equivalent’ to 199 keystrokes (25 lines times 7 keystrokes per line plus 24 keystrokes to enter new lines). According to the Chaitin-Kolmogorov-Solomonoff (CKS) idea of randomness, the first sequence has a low complexity whereas the second has a high complexity. By defining complexity as such, CKS established levels or degrees of randomness based on the shortest program that would describe a sequence. A sequence is ultimately random (often called Kolmogorov random) if the length of the program needed to produce the sequence is equal to the length of the sequence. Such a definition of randomness is different than that of Von Mises because, rather than requiring the randomness of a sequence to be judged by absolute unpredictability, it requires unpredictability by a small set of simple rules. In other words, every sequence that is not periodic, and hence cannot be described by simple rules, is random with some degree of randomness.

There is more to the CKS definition of randomness. Let us consider the following example. A sequence of 0’s and 1’s of length \( n = 3 \) can be represented as one of the following \( 2^3 \) possibilities:

\[ 010, 000, 111, 011, 110, 101, 001, 100 \]
There are several possible programs, which produce the above eight sequences of 0s and 1s. These programs can be programs of length-one (i.e. it requires only one line), or length-two (it requires two lines), or length-three. For example, the program of length-one,

\[
\text{write 1 three times}
\]

generates the sequence 111. Similarly, the program of length-one,

\[
\text{write 0 three times}
\]

generates the sequence 000.

There are no more length-one programs that will generate any of the remaining six sequences in our example. There are, however, four length-two programs that will generate four of the remaining six sequences. They are:

\[
\text{Step 1: write 0 twice}
\text{Step 2: write 1 } \} \text{ generates 001}
\text{Step 1: write 1 twice}
\text{Step 2: write 0 } \} \text{ generates 110}
\text{Step 1: write 0}
\text{Step 2: write 1 twice } \} \text{ generates 011}
\text{Step 1: write 1}
\text{Step 2: write 0 twice } \} \text{ generates 100}
\]

Finally there are two length-three programs that will generate the last two sequences:

\[
\text{Step 1: write 0}
\text{Step 2: write 1 } \} \text{ generates 010}
\text{Step 3: write 0}
\text{Step 1: write 1}
\text{Step 2: write 0 } \} \text{ generates 101}
\text{Step 1: write 0}
\text{Step 2: write 1 twice } \} \text{ generates 012}
\text{Step 1: write 1}
\text{Step 2: write 0 twice } \} \text{ generates 102}
\]

The last two programs are of the same length as the sequence they produce. Thus, given the eight possible length-three sequences in our example, the proportion that can be described by a program having less complexity than the sequences is 6/8. Accordingly, in this example, two of the eight sequences are of maximum complexity or truly random. As \( n \) increases, the proportion of length-\( n \) sequences that can be described by a program with less complexity (and thus being not maximally random) decreases and the proportion of the truly random sequences increases. For example, if \( n = 5 \) there are \( 2^5 \) possibilities but only \( 1/16 \) of these can be described by a less complex program. Since every number can be expressed in binary form, this example demonstrates that as \( n \) increases there will be infinitely more maximally random numbers than not. Consequently, we are far more likely to find randomness than order in the domain of mathematics! Therefore, we conclude that while there are plenty of random numbers we simply cannot prove that a given number is random. These conclusions form the basis of the so-called Chaitin’s theorem, which reminds us of Gödel’s theorem:

\[
\text{There exist numbers having complexity greater than any mathematical system can prove.}
\]

Imagine next, that in a room there are 10 chairs numbered 1 to 10. Also assume that in that room 10 individuals are standing each wearing a T-shirt with a different number on it. At the sound of a bell they select a chair and sit down. If the 10 numbers on the T-shirts are 2, 7, 9, 11, 20, 22, 23, 29, 30, and 31, one possible sitting arrangement could be the following:

\[
\text{Chair number: } 1 \, 2 \, 3 \, 4 \, 5 \, 6 \, 7 \, 8 \, 9 \, 10
\text{T-shirt number: } 7 \, 20 \, 22 \, 30 \, 23 \, 11 \, 29 \, 31 \, 9
\]

This is called a perfect sitting because at the end everybody is sitting and all chairs are occupied. In mathematical terms we say that the set of the chair numbers \{1, 2, 3, 4, 5, 6, 7, 8, 9, 10\} is of the same size as the set of the T-shirt numbers \{2, 7, 9, 11, 20, 22, 23, 29, 30, 31\}. The above two sets are both finite sets because they include a finite number of elements (10).

Now let us move to infinite sets. An obvious such set is the one containing all whole numbers:

\[
\{0, 1, 2, 3, 4, 5, 6, \ldots\}
\]

Similarly the set containing all positive integers (counting numbers)

\[
\{1, 2, 3, 4, 5, 6, 7, \ldots\}
\]

is infinite simply because infinity has no end and thus there must be infinite positive integers. Between these two sets we can thus make the following perfect ‘sitting’ assignment where each element of one set ‘sits’ upon one element of the other set.

\[
0 \, 1 \, 2 \, 3 \, 4 \, 5 \, 6 \, 7 \, 8 \, 9 \, \ldots
\downarrow \downarrow \downarrow \downarrow \downarrow \downarrow \downarrow \downarrow \downarrow \downarrow
1 \, 2 \, 3 \, 4 \, 5 \, 6 \, 7 \, 8 \, 9 \, 10 \, \ldots
\]

It follows that the two sets are of the same size, even though the second set is identical to the first set minus one element. This is a property of infinity. You can subtract something from it or you can add something to it and you still get infinity. Similarly, we can show
that the even numbers, the odd numbers, as well as the set of rational numbers, are all sets of the same size: the size of the counting numbers. Because of that all these sets are called countable.

Let us next assume that we have a list of all real numbers (rational and irrational) between 0 and 1, which shows each real number and its seat number. Let us also assume that all real numbers are given by some law or rule. Say that this list is as follows:

<table>
<thead>
<tr>
<th>Seat</th>
<th>Real number</th>
<th>Change bold digit to</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>0.813528 ...</td>
<td>3</td>
</tr>
<tr>
<td>2</td>
<td>0.772134 ...</td>
<td>4</td>
</tr>
<tr>
<td>3</td>
<td>0.023577 ...</td>
<td>9</td>
</tr>
<tr>
<td>4</td>
<td>0.312153 ...</td>
<td>0</td>
</tr>
<tr>
<td>5</td>
<td>0.552728 ...</td>
<td>6</td>
</tr>
<tr>
<td>6</td>
<td>0.856571 ...</td>
<td>7</td>
</tr>
</tbody>
</table>

Take a look at the first digit in the expansion of the first number. It is 8. Let us now choose a number that is different than 8 and put it aside. We can do this by putting the digits 0, 1, 2, 3, 4, 5, 6, 7, 9 inside a box and select one blindly. Say this number is 3. Now consider the second digit in the expansion of the second number. It is 7. Do the same trick and choose a number that is not 7 and put it aside. Say this number is 4. Continue this process with the third digit of the third number, the fourth digit of the fourth number, and so on. The following is a possible summary of this process:

<table>
<thead>
<tr>
<th>Seat</th>
<th>Real number</th>
<th>Change bold digit to</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>0.813528 ...</td>
<td>3</td>
</tr>
<tr>
<td>2</td>
<td>0.772134 ...</td>
<td>4</td>
</tr>
<tr>
<td>3</td>
<td>0.023577 ...</td>
<td>9</td>
</tr>
<tr>
<td>4</td>
<td>0.312153 ...</td>
<td>0</td>
</tr>
<tr>
<td>5</td>
<td>0.552728 ...</td>
<td>6</td>
</tr>
<tr>
<td>6</td>
<td>0.856571 ...</td>
<td>7</td>
</tr>
</tbody>
</table>

We can now form a number between 0 and 1 whose decimal expansion includes all the digits in the third column in the above order. This number is

0.349067 ...

This number is different from the first number in our example because the first digits in their expansion do not match. It is different from the second number because their second digits do not match. It is different from the third number because their third digits do not match, and so on. Going down the list this way, effectively creates a number that is different from all numbers in the list. This means that it cannot be in the list! But it is a real number and we started with the assumption that our list includes all real numbers. Since whatever we did was mathematically legal, this contradiction implies one thing only: in the case of real numbers the perfect seating does not exist. Therefore the set of real numbers is uncountable. In fact we see that the set of real numbers is bigger than the set of the rational numbers or the set of whole numbers. Since all these countable sets are infinite sets this result indicates that the set of real numbers is a bigger infinity. It follows that there is more than one infinity. In fact, one can extend the above examples to create an infinite number of infinities (the transfinite numbers as they are called), which become successively larger in size up to the ‘ultimate’ infinity that engulfs all other infinities. These insights are largely due to the works of Georg Cantor, a Russian born mathematician.

The conclusion that the set of real numbers is not countable is far reaching. One may ask the question: where do all these ‘unaccounted’, ‘extra’ numbers come from? When we started our discussion on the size of the set of the real numbers we assumed that we had a list of all real numbers and that all of them are given by some kind of a rule. But when we produced the real number that was not in the list we actually picked the digits in its expansion at random. No particular rule was applied in choosing its digits. They were chosen at random. Then these ‘extra’ numbers are the product of randomness. Interestingly, when we deal with infinity, mathematics cannot do away with randomness. Thus, the discussion in this section gives a new perspective in the definition of randomness. Since sequences exist that obey no rule, Von Mises’ definition of randomness may after all be a valid one. Similarly, the interpretation of Chaitin’s theorem becomes clearer; maximally random numbers cannot be described in any mathematical system, meaning that randomness does not manifest itself as a set of rules but as the absence of rules. The construction of the number 0.349067... clearly demonstrates that there is no deterministic algorithm that will produce this sequence and thus we cannot possibly predict the next digit by antecedent digits. Thus, pure randomness implies unpredictability. This point is the central point behind the definition and sources of randomness.

### Randomness of the first kind

Imagine that we are given or we observe a pattern-less sequence that has been generated by some rule. If we are not able to extract the rule, what is the difference between such a sequence and a truly random sequence? Many will argue that for all practical purposes there is no difference. Thus, our inability to extract the rules
makes predicting future digits impossible and thus constitutes a source of randomness. Here, however, we have to be careful with what we mean by ‘our inability to extract the rules.’ Do we mean that in principle we could get to the rules but we do not have the knowledge to get to them or that there is no possible way to get to the rules? If the former is true then strictly speaking the sequence is not random. It is just waiting to be ‘debugged’. In this case the rules are actually reversible and thus in principle we can go backwards and find the rules from the sequence. If the latter is true, then the procedure to construct the sequence is irreversible and thus, while we can go from the rules to the sequence, we cannot ever go from the sequence to the rules. It is this inability to recover the rules that will generate truly random sequences. But how does such irreversibility occur?

Consider again the first mathematical operation which, for simplicity we will call operation \( O_1 \): Start with a number. If this number is even, multiply it by 3/2; otherwise add 1 and then multiply by 3/2. This generates the second number. Repeat the above step to produce the third number and so on. Operation \( O_1 \) results in a sequence of odd and even numbers, which we will denote by \( S_1 \). Once we have \( S_1 \) we apply the second operation, which we will call operation \( O_2 \), according to which an odd number is replaced with 1 and an even number with 0. This leaves us with a sequence of 0’s and 1’s, which we will denote as sequence \( S_2 \). If the starting number was the number 1, then operation \( O_1 \) produces the following sequence \( S_1 \) of odd and even numbers:

\[
S_1: 1, 3, 6, 9, 15, 24, 36, 54, 81, 123, 186, 279, 630, 945, 1419, 2130, 3195, 4794, \ldots
\]

Subsequently, operation \( O_2 \) produces the following sequence \( S_2 \):

\[
S_2: 1101100011010110101\ldots
\]

The above two operations define a mathematical system, which is isolated from external influences and in which the initial condition is simple and well defined. Sequence \( S_2 \) does not appear to have a coherent pattern (this becomes even more apparent if we continue the process for many steps). However, since it is generated according to a set of rules, it is not random. Given the rules, any future value can be calculated or predicted. There is, however, a catch here. If you were given part of \( S_2 \) but not the rules that generated it, would you be able to predict the next digit? The answer to this question is no and Fig. 3 explains why.

Given sequence \( S_1 \), a patient person might at some point figure out operation \( O_1 \). In this case we can go from operation \( O_1 \) to sequence \( S_1 \) and vice versa (this reversibility is indicated by the bidirectional arrow between operation \( O_1 \) and sequence \( S_1 \)). Similarly, one might speculate that the zeros and ones represent even and odd numbers, respectively. This will allow us to go from operation \( O_2 \) to sequence \( S_2 \) and vice versa. However, once we have figured out operation \( O_2 \), there is no possible way to go to sequence \( S_1 \) because for each zero there is an infinite number of even numbers and for each one there is an infinite number of odd numbers to choose from and the initial number can be anything. This irreversibility is indicated by the unidirectional arrow between sequence \( S_1 \) and operation \( O_2 \). We thus see that operation \( O_2 \) injects an uncertainty that inhibits us from recovering the rules of the construction and makes \( S_2 \) maximally random. While not all sequence constructions may be irreversible, the above example clearly demonstrates that extracting underlying rules may not always be attainable.

The loss of information, which results from this irreversibility, is not just a well-thought-out mathematical trick. Information initially contained in a system could indeed get lost during its evolution. Imagine two compartments separated by a diaphragm, one filled with water at 40° F and the other with water at 90° F. If the diaphragm is removed, the two water samples mix and produce a sample, which is at a uniform temperature throughout. In this final sample all the information about the initial temperatures is lost and cannot possibly recovered no matter how knowledgeable we are.

Order and predictability arise from rules. Randomness and unpredictability arise from the absence of rules. This source of randomness is, however, ideal if not trivial. In the mathematical system
and in the physical world there is always some kind of an underlying rule(s). Here we see that unpredictability and thus randomness may arise from irreversible programs or procedures, which inhibit us from getting to the rules not just from the absence of rules. We will call this randomness, \textit{randomness of the first kind}.

\textbf{Randomness of the second kind}

Consider an \textit{iterative process} where the same operation is repeated, every time using the result of the previous step as the starting point. The most famous such process is given by the logistic equation $x_{n+1} = 4x_n(1 - x_n)$, which has been used to study population dynamics. The number 4 is the only parameter of this equation. Since this system is described by only one equation it is obviously a very simple system. For $x = 0$ or $x = 1$ all subsequent values become zero and for $x > 1$ or $x < -1$ the population becomes negative. Thus, for nontrivial dynamics (i.e. requiring that the population does not become extinct or negative) the values of $x$ must range between 0 and 1. Iterating the logistic equation from an initial value of 0.4 results in the evolution shown in Fig. 4.

What we observe is that $x$ goes up and down in an apparently irregular way. No apparent pattern is evident. For all practical purposes this signal is random. Nevertheless, it was generated from a well-defined initial condition (0.4) and a very simple rule.

Figure 5 shows $x$ as a function of the time step for initial condition 0.4 as well as for initial condition 0.7. We observe that the two evolutions are different. They are aperiodic and they do not converge to the same result. Somehow the system remembers its initial condition forever. The evolution of this simple system is clearly dependent on the initial condition. It gets even more interesting. Let us assume that the second initial condition was not 0.7 but was 0.405 (about 1% different from the first initial condition). If we compare the two evolutions again (Fig. 6), we see that they start very close to each other, but they soon diverge and follow completely different paths. Not only is the system sensitive to the initial condition, it is sensitive to even the tiniest of fluctuations. And not only does the system not forget a tiny fluctuation, it actually amplifies it and soon the two evolutions diverge significantly.

As we have seen earlier, aperiodicity does not necessarily mean unpredictability. For example, the $\pi$ digits are aperiodic but we can predict any digit we want. But how about sensitivity to the initial conditions? Could sensitivity to initial conditions create randomness and

\begin{figure}[h]
\centering
\includegraphics[width=0.4\textwidth]{fig4.png}
\caption{Solution of the logistic equation from the initial condition 0.4.}
\end{figure}

\begin{figure}[h]
\centering
\includegraphics[width=0.4\textwidth]{fig5.png}
\caption{Solution of the logistic equation from the initial conditions 0.4 and 0.7.}
\end{figure}

\begin{figure}[h]
\centering
\includegraphics[width=0.4\textwidth]{fig6.png}
\caption{Solution of the logistic equation from the initial conditions 0.4 and 0.405.}
\end{figure}
make a system unpredictable? From the above example one can forcefully argue that the initial condition can be specified exactly. For example, we may specify 0.4 as our initial condition. Then the equation will produce all future values. Thus, one may argue, sensitivity to the initial conditions is not a condition for unpredictability. This is a strong argument but there is a little problem with it.

The five first values of the evolution of the equation \( x_{n+1} = 4x_n(1 - x_n) \) from the initial condition \( x_0 = 0.4 \) are:

\[
\begin{align*}
    x_0 &= 0.4 \\
    x_1 &= 0.96 \\
    x_2 &= 0.1536 \\
    x_3 &= 0.52002816 \\
    x_4 &= 0.9983954912280576 \\
    x_5 &= 0.0064077372941726356570612432896 \\
\end{align*}
\]

What do you observe? The digits after the decimal point increase (in fact double) with every iteration. After seven iterations the result carries 128 digits. After 12 iterations there are 2048 digits! The number of digits is actually given by \( 2^n \), where \( n \) is the number of iterations. Calculating exactly out to only 100 steps will require a computer that will carry calculations with \( 2^{100} \) decimal points. This number is approximately equal to \( 10^{30} \), which is one-trillion-times greater than the age of the Universe in seconds. Computers do not routinely handle more than a hundred digits. So what does the computer do when the iteration reaches the point where the digits are more than the digits the computer can carry? It simply rounds off the result or chops off the extra digits. That in effect makes the result an approximation to what the result would have been if the computer had the ability to carry calculations with unlimited number of digits. This approximation will now play the role of a fluctuation, which will be amplified and soon lead to an evolution that will be completely different than the actual one. Thus, only if we have infinite precision and infinite power we will be able to predict such systems accurately. Because we do not have that, for systems that are sensitive to the initial conditions, the exact state of the system after a short time cannot be known. The outcome of such systems after that time is simply random as small fluctuations amplify enough to dominate the evolution of the system. Thus, a future state is unpredictable. Note that the logistic equation is what we call a non-linear equation.

We thus see that randomness is created even if we know the rules. In this case the source for the randomness is not rule irreversibility or our inability to find the rules but our inability to have infinite precision and infinite power. This kind of randomness has been termed chaos and it is distinctly different from the first kind. Chaos is strictly a property of nonlinear systems.

**Randomness of the third kind**

Now let us consider the following example. The distance between your home and the shopping mall is fixed. It is always the same, never varies. We all know that if the speed of an object is constant, then the time that it will take to go from A to B is equal to the distance between A and B divided by the speed. It follows that as long as the speed does not change, the time that it takes to travel the distance AB will always be the same. Now assume that you travel with your car (our ‘system’ in this example) from your home (A) to the mall (B). You know that the distance is 5 miles so you figure that at a constant speed of 50 miles per hour it will take you one-tenth of an hour or 6 min to reach the mall. Would you bet money on such a prediction? I hope not, because it will never be exactly 6 min. A slower driver in front of you, a driver that suddenly decides to ‘cut’ in front of you, a yellow traffic light that forces you to decide whether to stop or accelerate, the presence of a police car, the sound of a honk, and many other ‘external’ factors will cause you to depart from the constant speed of 50 miles per hour. Since the number of these factors is not fixed, each time you go to the mall it will take a different length of time. This makes the actual length of the trip very unpredictable, which means that the duration of the trip is a random number. Thus, even though we know the rules of the system (so there is no randomness of the first kind) and we can assume that our system is not chaotic (which excludes randomness due to sensitivity to the initial conditions) we still end up with randomness.

The above example introduces us to a mechanism for randomness, where randomness is explicitly introduced into the underlying rules of the system. It corresponds to saying that there is some kind of external environmental component with essentially uncountable ‘agents’ continually affecting the system with their actions. Such processes are called stochastic processes. The word stochastic comes from the Greek word στοχαστικής, which refers to the person who learns about future events or hidden things by means not based on reason.

Our ‘system’ here may be thought of as a mathematical system described by a set of simple rules or equations whose evolution can, in principle, be computed.
by solving the equations. When the system is exposed to external influences, however, its behavior is modified. But what is this external ‘noise’? Where is it coming from? In our example, the environment is represented by the other drivers and the traffic lights system, which can also be thought of as simple systems. In this case then, what we have is many simple systems interacting. Each system is very simple but the collective behavior of many interacting systems may be very complicated.

It is then logical to assume that in our example the ‘system’ is a subsystem of a grand system (possibly the Universe) where many subsystems operate according to their rules and interact between each other. As subsystems interact they exchange information. Information received by one subsystem from another may interfere with its rules, thus producing an unexpected result. Such interactions, especially in a large number of subsystems, create an extremely complex behavior that can only be studied using probability theory. This ‘stochasticity’ is our third kind of randomness: randomness generated by the continuing effects of the environment.

It is interesting to note here that under this scenario very complicated behavior and randomness can be generated even if we start with no randomness of any kind. The theory of nonlinear dynamical systems has clearly established, for example, that many systems, which exhibit a very regular (periodic) behavior, become irregular and aperiodic when they are coupled with an external force, which is also very regular. This may also lead to randomness of the first kind or to chaos. Whatever the case might be, one thing is certain: very simple rules can either alone or in combination with other simple rules create randomness and unpredictability. And because we are often dealing with an infinite number of interacting ‘agents’, and the only way to study and predict their collective behavior is stochastically.

It is also interesting to make some connections with real life at this point. In life, accidents (randomness) happen (1) when we do not know how things work (we do not know the rules), (2) when we do not pay full attention (we do not compute accurately), and (3) when we interact with people who affect our lives. Doesn’t this look like randomness of the first, of the second, and of the third kinds, respectively?

Finally, we should keep in mind that behind all mechanisms of randomness lurks the notion of infinity. Whether it is the absence of infinite knowledge or infinite power or the interplay of infinite agents, one cannot avoid infinity. It is the arena where the interplay of rules and randomness takes place. If this arena disappears, all evolutions are doomed to repeat. For example, it has been shown that because the computer can only carry a finite number of digits in its calculations, the round-off error will force a non-periodic trajectory to coincide with a point in the past rather than simply coming very close to it. Once this occurs, the evolution has no choice but to repeat (Tsonis 1991). Consider again the first five exact values that result when we start iterating the logistic equation from an initial condition of 0.4:

\[ x_0 = 0.4 \]
\[ x_1 = 0.96 \]
\[ x_2 = 0.1536 \]
\[ x_3 = 0.52002816 \]
\[ x_4 = 0.9983954912280576 \]
\[ x_5 = 0.00640773729417263956570612432896 \]

If we assume that the calculator or computer used to perform these calculations can only carry one decimal point then the first value will be truncated to 0.9. If the value of 0.9 is used rather than the actual value of 0.96 to calculate the second iterate we get a new value of 0.36 which will be truncated to 0.3. Continuing like this we will find that the fifth iterate is truncated to 0.9, which is the value of the first iterate. From this point on we simply start over and the evolution becomes periodic.

Part 2: Randomness in the Universe

Quantum mechanics

The fact that light has properties of waves was known since the time of the British physicist Thomas Young who, as early as 1801, performed the two-slit experiment demonstrating that light like waves in the sea create diffraction and interference patterns. However, almost 100 years later, Einstein observed that when light falls on a metal plate, the plate ejects a shower of electrons. He further observed that the shorter (the more energetic) the wave the higher the speed with which the electrons are ejected. This is not what would happen if light were a wave. This is more like what happens when two particles collide.

Thus, light appears to behave as both particle and wave. This establishes the wave–particle duality of light and proves that electromagnetic waves can behave as particles. But can particles behave as waves? This question was posed by the French physicist de Broglie who suggested that electrons, particles of a certain mass, could be treated as systems of superposed waves or as wave packets. Wave packets do not just describe pure waves or just pure particles...
but a combination of both. With such a description, de Broglie was able to study the quantum-mechanical motion of a particle and to predict the magnitude of the wavelength of an electron. And here lies the beauty of scientific ingenuity and reality. Scientists can always propose a theory which, based on mathematics, will make predictions. But it is not until the predictions are verified that a theory becomes accepted. In the case of de Broglie’s hypothesis, it was not long before experiments not only verified that electrons have characteristics of waves (they create interference patterns) but also recovered the predicted wavelength of the electron. Subsequent experiments went even further to show that even larger particles, such as molecules, behave as waves and that their wavelengths are exactly those predicted by de Broglie’s theory. Soon after that the Viennese physicist Schrödinger developed the equation of motion of a particle whose solutions were the de Broglie waves. Before long, in 1927 the German physicist Werner Heisenberg starting from the hypothesis that an electron is a wave packet obeying a wave function proved his famous uncertainty principle, which ever since has made quantum mechanics as mystic a science as we will ever have.

In classical mechanics what specifies the complete state of a particle is its velocity and position. If we know these two variables we can solve the equations of motion and predict the position of the particle at any time in the future. Such complete determinism is the most fundamental aspect of classical mechanics. The uncertainty principle states that when it comes to subatomic particles, such as photons and electrons, one cannot measure exactly the position and velocity of a particle. More specifically it says that measuring the position with high accuracy results in a great uncertainty in the value of the velocity and vice versa. Consequently, we cannot ascertain the exact position of a particle without losing information about the velocity and vice versa. In quantum mechanics we cannot have well-defined states for the position and velocity; we can only have a quantum state, which is a combination of position and velocity. Since the position and velocity can only be known approximately, this state is defined by probabilities of the position and velocity. For example, “the particle’s position is most probably somewhere there and its velocity is most probably around that value.” Thus, quantum mechanics introduces an element of unpredictability or randomness in the scientific description of subatomic particles. Our Universe emerges as intrinsically non-deterministic and unpredictable.

Imagine an isolated room in a void and a coin spinning on a frictionless table. With such a setup there is nothing to interfere with the spinning of the coin, and because the table has no friction there is nothing to slow down the spinning of the coin. The coin spins endlessly and will never tumble on the table. In this case both states of the coin (heads and tails) are superposed. None of them will be observed. If, however, an observer interferes by trying to make some measurement about the coin, any tiny fluctuation will cause the coin to slowly tumble on the table. This will then yield an outcome: either heads or tails. Light has the property of the ever-spinning coin. Both possibilities, or states, are superposed. The particle is allowed to do everything possible simultaneously. It is a particle and a wave at the same time. And in a similar way, in experiments designed to observe the properties of light, scientists concluded that the observer interferes and what is observed is either a particle property or a wave property. And here comes the mystic part. These conclusions postulate that the superposition of states collapses into a classical outcome (either particle or wave but not both) once an observation is made. It follows, that an outcome of an observation is relevant to the observation and thus not objective. As such, in the quantum realm, this postulate introduced a strange connection between the observer, measurement, and physical reality, which soon runs into trouble. Einstein used to make fun of the scientists who proposed this explanation (the Copenhagen interpretation) by asking them whether they believed that the moon existed only if they looked at it.

To do away with the troubles of the Copenhagen interpretation, in the 1950s the American physicist Hugh Everett III proposed another interpretation. He suggested that when an observer checks to see if an electron is a particle or a wave, the Universe splits into two universes. In one the observer sees a particle and in the other he sees a wave. In other words, both outcomes happen but they happen in different universes. If you think of the electron’s particle–wave duality as a coin, Everett’s interpretation implies that when we toss a coin we do not get just tails or just heads but we get tails and heads at once but in different universes. This is the many-worlds interpretation. According to this interpretation when the phone is ringing while you are reading a book, quantum effects in your brain lead to a superposition of outcomes (let the answering machine get a message and keep on reading, stop reading and answer the phone, ask your daughter to answer the phone, and so on), all of which happen in different universes. Thus, random quantum processes cause the universe to split into as many copies as the possible
outcomes. Each of these copies of you is unaware of the others. The branching is noticed only as a slight randomness in your decision. The next time we need to make a decision each universe splits again and so on to basically create an unimaginable number of universes, which describe all possibilities that there ever were. Crazy enough? Well, the reactions to this theory have ranged from ridiculous to ingenious. Many argue that this is very wasteful and uneconomical of nature. Others counter argue by asking, "Exactly what is nature wasting? Certainly not space or mass. There is an infinite amount of both to begin with."

There is no reason to believe that our state of knowledge of the mysteries of the Universe is complete or even nearly complete. If Newton had some evidence that time is not absolute, his physics would not have been able to explain it. To him it would appear bizarre and weird that two observers will measure different times between two events. Similarly, it may be that the bizarre interpretations of quantum mechanics appear bizarre because some fundamental insight of our Universe is yet to be understood or discovered. However, as it does not mean that Newtonian physics is all wrong just because of the unexplained evidence of the relativity of time, similarly it does not mean that quantum mechanics is wrong because the explanation of the observed randomness at small scales appears bizarre.

**Chaos**

Above, when we discussed randomness of the second kind, we demonstrated the sensitivity to the initial conditions and the definition of chaos using a simple mathematical equation that does not have a direct relation to a physical problem. This property, however, is not just the property of a mathematical system. It is found in natural systems as well. Back in the early 17th century, the German astronomer Johannes Kepler published his first law, which stated that the orbit of an object around an attracting body is an ellipse with the attracting body located at one of the foci. The ellipse remains constant in space, the speed, however, of the orbiting body varies. According to Newton’s gravitational law, the force of attraction is proportional to the product of the masses of the two objects and inversely proportional to the square of their distance. Since the orbit is an ellipse the distance between the two bodies is not the same at all times. As such the gravitational pull varies; it is greatest at the pericenter and smallest at the apocenter. From Newton’s second law it then follows that the speed of the orbiting object varies accordingly. Nevertheless, the position and speed of the orbiting object are determined at any time and they are regular. They repeat exactly after a fixed time interval.

The situation, however, becomes a bit more complicated when there are more than two bodies in the picture. For example, Earth attracts the moon while both are attracted by the Sun. What is the motion of the moon in this case? The problem can be exactly described by a set of nonlinear equations. However, the problem has no analytic solution. In other words we are not yet able to find a solution using standard mathematical approaches. The only way to solve such problems is numerically. If the calculations are done with sufficient numerical accuracy and for short time intervals, we can track the motion of the objects for a long time. This procedure can today be done efficiently with a computer. At the time of Kepler and Newton, however, this was not possible and both of them, while aware of the problem, saw this irregularity as a nuisance. It was not until the early 20th century when the French multi-scientist Henri Poincaré showed that the numerical solution to the three-body problem is very irregular and very sensitive to the initial condition. In fact Poincaré discovered chaos but due to the unavailability of computers he could not study this problem in detail. In 1925 a glimpse into the complexity of this problem was provided by a computation carried out by 56 scientists under Elis Stromgren at the Observatory of Copenhagen, which showed a solution to the so-called restricted three-body problem, which deals with the orbit of a moon under the gravitational influence of two planets. This work, which was published in 1925, took 15 years to complete (due to the lack of computers). Today such a computation would take a few hours on a desktop computer. Nevertheless, for the first time it was realized that irregular behavior can be observed in a very simple system that describes a natural phenomenon and that sensitivity to the initial conditions will make the behavior of the system practically unpredictable. This is the same property we discussed with the logistic equation, which we termed chaos.

The theory of chaos had to wait several decades until the development of fast computers allowed such calculations. Then, in 1963, Edward Lorenz, an atmospheric scientist at MIT, who was trying to explain why weather is unpredictable, reduced the complicated physics of the atmospheric circulation into three simple nonlinear differential equations (a differential equation describes changes of a variable in time), which modeled the behavior of a fluid layer heated from below. This is an approximation of what happens basically every day in the lower atmosphere of our planet. The Sun rises,
the surface of planet absorbs solar radiation and gets warm. Subsequently, the air gets warm by contact with the warmer surface and rises. This rising motion leads to turbulent motion. When Lorenz solved the equations and plotted the results, he was surprised to see that this turbulent motion was behaving quite randomly and never repeating exactly. In addition, Lorenz found that this system is sensitive to the initial conditions. As with the logistic equation, evolutions from two slightly different initial conditions soon diverge and follow different evolutions. If we think of these two slightly different initial conditions as the true state of the atmosphere and what we actually measure (measurements always include some error, so we never really measure the true state), then their divergence indicates loss of predictability. This was the first time that somebody provided a scientific reason for why weather cannot be predicted with accuracy after a few days. Lorenz published his results in the highly respectable Journal of Atmospheric Sciences (Lorenz 1963). At that time, however, meteorologists were occupied with other problems and did not pay attention to this remarkable paper. It took more than a decade before mathematicians and physicists discovered the paper and for the theory of chaos to take off and develop into one of the most important scientific theories of the 20th century.

Does this mean that chaos is a major property of our Universe?

The problem with answering this question is that while individual systems might be chaotic, observations are often the result of many systems (some of which are chaotic some of which regular) interacting and affecting each other. For example, when we measure the vertical speed of the air inside a cloud, what system do we probe? Is our system the cloud itself? Or is it the atmosphere or maybe the Earth or even the solar system? As we discussed above, in this case the actual chaos may be masked and what we get is stochasticity. Nevertheless, in the last three decades laboratory experiments as well as measurements of natural processes have shown that many phenomena in many areas of science are chaotic. This evidence makes sensitivity to the initial conditions a fundamental property of nature.

Chaos has been called by the late American physicist Joseph Ford ‘Gödel’s child’. Just as Gödel’s theorem tells us that there will always be questions in any particular logical system that cannot be answered, chaos tells us that there are physical questions that cannot be answered like, what the weather is going to be in New York on 19 November 2021 (or some other date far in the future). For such a prediction, the initial state of each molecule in the atmosphere worldwide should be available to a precision that exceeds the limits imposed by quantum theory.

We should mention here that the unpredictability associated with chaos in natural systems, like weather for example, is more complicated than that of an abstract mathematical system, like the logistic equations, where the initial condition can be specified exactly. In natural systems we have to measure the initial condition. For example, to make a weather prediction we measure the temperature, pressure, moisture, and other variables and then we set the system (equations) in motion and see what happens. Measurements as we all know are subject to errors. Every time I travel to my office, I pass places where digital thermometers show the temperature. Somehow they all differ. This may be due to the natural variability of the temperature field, but I have noticed that the two thermometers in my house also never agree. Instruments are simply not exact. This will cause an uncertainty in the measurements used to specify the initial condition of the atmosphere. In this case we will start with an error. That error will couple with the round-off error introduced by the computer and things will go bad even faster. Not to mention that the initial state of the atmosphere is measured only at certain locations, thereby missing a lot of information in between. This results not only in an inexact initial condition but also in an incomplete one.

**The supreme law**

Imagine a container with two compartments, A and B, separated by a partition and isolated from its environment. Also imagine that A is full of air while B is empty. If we remove the partition what will happen? Obviously the air will expand to occupy both compartments. The opposite phenomenon where the air in a room suddenly accumulates in one half leaving the other half empty never happens. The impossibility of such events is due to the second law of thermodynamics, often hailed as the supreme law of nature.

Let us consider in more detail the example with the two compartments. Before the partition is removed the particles that make up the air are all in compartment A. This picture actually represents an ordered state, simply because there are restrictions for the particles. When the partition is removed the particles have no restrictions and they can move anywhere they wish. Eventually they uniformly occupy both compartments.

This is a state of equilibrium and from that point on, even though the particles are free to move all over,
chances are that the same number of particles will be found in A and in B. This equilibrium state represents a state of lower order or higher disorder. Now would you call this process a reversible process? In other words would you expect that the particles would just by themselves return to the original ordered state? Some will argue that based on probability theory there is always a chance that somehow all the particles will be found in compartment A again, but I would not bet money on this. Even with a limited number of particles this may take the age of the Universe before it happens. We can, therefore, safely assume that this process is irreversible. This irreversibility is directly related to the increase of disorder. Physically, this expresses the second law, which says that during an irreversible process the entropy of an isolated system increases. Here, just to be on the safe side, we must mention that our system of the two compartments is considered isolated. In other words, it is alone and not interacting with anything else. In this case we cannot argue that an external force can be invoked, which will physically move all particles back to compartment A, thereby decreasing the entropy. By the way, entropy comes from the Greek word ἐντροπία, which means the ‘inner behavior’.

In nature processes tend to be irreversible. In fact, unless a process is very much controlled by an experiment it is always an irreversible one. A cloud forms, it rains, and then dies out. You never see the opposite. A cup falls and breaks. You never see broken fragments rising and forming a cup. The interpretation of this law for the fate of our universe is fairly straightforward. Assuming that our universe is an isolated system, all the transformations that happen within it result in a steady increase of entropy with time. As such the Universe is evolving toward a state of maximum entropy. Therefore, once the maximum entropy has been reached it cannot be increased any further. That simply means that there cannot be any changes any more. Thus, the maximum entropy corresponds to equilibrium and the second law describes the general tendency of the Universe to reach equilibrium.

All this sounds very matter of fact. Unlike quantum mechanics, nothing is weird about our discussion in this section. You may be wondering why, in this case, there is no place for randomness. Let us consider that we are dealing with four particles all of them being initially in compartment A. In how many different ways can we arrange four particles in A and no particles in B? The answer is obviously in one way. Four particles in A and zero in B. Now we know that if we remove the partition the particles will move around and will occupy both compartments uniformly (i.e. without a preference for either compartment). Thus, we would expect that, at any time, two particles will be in A and two in B. In this case, in how many different ways can we have two particles in A and two particles in B? The answer to this question is six.

The number of different ways to arrange particles in the two compartments is called the number of complexions. Thus, we see that during our irreversible process of particle dispersion, where the entropy is increasing, the number of complexions is also increasing. In the late 1800s the Austrian physicist Ludwig Boltzmann proved that this number of complexions is directly related to the entropy of the system and indeed as the number of complexions increases entropy increases proportionally.

Are the above possibilities the only ones? Could we not have three particles in A and one in B or vice versa? Of course we can. In this case the number of complexions is four. Thus, altogether (including the possibility that all particles are in B) there are 12 possible particle configurations. Six out of those 12 configurations correspond to maximum disorder, or maximum entropy, two correspond to minimum entropy and four to some intermediate value.

It follows that the most probable state is the state of maximum entropy. It also follows that the irreversibility of natural changes does not result from certainty but from probability. There is a higher probability to tend toward the state of maximum entropy than otherwise. We have thus discovered that the essential way in which systems evolve is statistical and that in nature irreversibility is associated with randomness. And since, from the small-scale order we go to the large-scale disorder, unlike in quantum mechanics where randomness defines the micro-cosmos, the second law tells us that probability rules the macro-cosmos. For completeness, Boltzmann’s mathematical relationship between entropy (S) and the number of complexions (P) is

\[ S = k \ln P + \text{constant} \]

where

\[ P = \frac{N!}{N_1!N_2!} \]

and \( N_1 \) is the number of particles in A and \( N_2 \) the number of particles in B.

Finally we should mention that since all subatomic particles obey quantum mechanics and all matter is made out of these particles, quantum mechanics is considered the most fundamental theory of nature. The rules of classical mechanics, which are followed by large objects (regular or chaotic), should somehow emerge from quantum mechanics. It is like an
impressionist painting, which though fuzzy when viewed close-up, produces a coherent picture when viewed from afar. The connection between quantum mechanics and the macro scale has not been achieved yet. This may be because quantum theory is not complete. Or it may be that we do not understand it completely. Nevertheless, quantum mechanics has seen many successes and applications in many areas. These include the prediction of the existence and subsequent discovery of the particle positron, the explanation of the formation of a positron and an electron when electromagnetic energy interacts with matter, the operation of transistors, and the development of lasers. Due to this, while it may be that it needs refining, quantum mechanics and its randomness is here to stay. Quantum mechanics, chaos, and the second law are prime examples that our Universe in its infinite space does not just obey rules but that it is also inherently random. Here again, as with the mathematical system, we find the concepts of infinity, randomness, and rules, interweaved and working together.

**Randomness of the fourth kind?**

Consider the game of chess. In chess 32 agents interact according to specific rules. For the masters, the game evolves according to a plan but there may be moments where for the next move more than one possibility may exist. Because of the limited time between moves, the player cannot possibly go through all the possible configurations and often has to use an ‘educated guess’ or his free will. Based on the choice, the outcome may be different. This uncertainty in the final result is randomness introduced by free will. A high-speed computer, on the other hand, could run all possible configurations and possibilities in the allotted time. In this case the computation is faster than the evolution of the game and there is perfect predictability: the computer will beat a human opponent all the time. As another example, imagine a soccer player leading an attack toward the opponents’ net. During his run, he often has to pass the ball to one of his teammates. The outcome of the attack depends on which teammate will get the ball. Our player can pass the ball to several of his teammates, but as he is advancing toward the opponents’ net he has little time to compute all the possibilities open to him. Here again the computation is slower than the evolution of the game and hence the player cannot make an accurate prediction. As is often the case, our player makes a choice, which even though he may be using his best judgment may not be the best choice. He simply uses his free will.

The issue of free will is rather controversial. By definition, free will is the conviction that humans have the capacity to choose their actions. It is a very divisive issue among philosophers and this author is the last scientist who will argue with philosophers.

However, in the scientific realm (which is of interest here), free will implies that decision making is not completely and necessarily determined by a physical prior cause(s). If we accept the view of a completely deterministic universe then there is simply no free will or randomness whatsoever. Everything has a cause and the only reason we do not understand or cannot predict is absence of complete knowledge. If on the other hand we reject determinism altogether, then everything that happens is independent of what happened before. What modern science and mathematics point to, however, is that both these two extremes are just extremes. The discussion in this paper presents plenty of evidence that, in our Universe and in the mathematical system that describes it, determinism and randomness coexist. I personally cannot but accept the fact that free will exists in humans and that free will choices do not necessarily require prior causes. In this case free will actions may through humans introduce randomness in the Universe. I will not go into the ‘weird’ topic of whether an electron has free will (I will leave this to philosophers) but it is easy to explain how free will emerges in a deterministic universe.

We can thus see that there could be instances when humans would inject randomness into their environment. The open question is whether or not it is only humans that have free will or that it is also a property of the Universe as a whole. Could it be, for example, that the Universe is a cellular automaton (as Stephen Wolfram contends,) and that it is impossible to devise a simulator that runs faster than it, thereby, causing free will to emerge? I will not attempt to get myself into such an abstract and philosophical issue but I will quote Ray Kurzweil from his ‘Reflections on Stephen Wolfram’s A New Kind of Science’, “… it should be noted that it is difficult to explore concepts such as free will and consciousness in a strictly scientific context because these are inherently first-person subjective phenomena, whereas science is inherently a third person objective enterprise. There is no such thing as the first person in science, so inevitably concepts such as free will and consciousness end up being meaningless. We can either view these first person concepts as mere illusions, as many scientists do, or we can view them as the appropriate province of philosophy, which seeks to expand beyond the objective framework of science.” (Kurzweil 2003).
After all the discussion so far, you may wonder why nature chooses to be irregular and unpredictable. Why is our Universe not simple and regular? Well, this is indeed an interesting question, which will be answered soon, but first we have to make connections between the sources of randomness in the mathematical system and those in the physical system.

**Part 3: Connections**

Figure 7 summarizes the possible connections between the mathematical and physical systems. First, let us take quantum mechanics. The randomness introduced by the uncertainty principle results from assuming that a photon or an electron is a system of superposed states or a wave packet. These wave packets make use of the particle–wave duality. We know that the duality is a fact. But we are still unable to explain why the duality exists. In this regard one may argue that we simply cannot recover at this point the rules and as such randomness of the first kind is generated. One might also conjecture that due to the inherited uncertainty in quantum mechanics, it will be subject to chaos. The area of quantum chaos is of great interest in science today. Indeed, there have been many indications that quantum systems display chaotic characteristics. This links quantum mechanics with randomness of the second kind. Now suppose that, as is implied by modern developments in theoretical physics, the universe began as a 10-dimensional bubble of space out of which only four (time and the three spatial) dimensions expanded to form the universe we live in. The other six dimensions simply compacted to form what we call subatomic particles. In a sense then these particles live in a 10-dimensional space, whereas they are observed in a four-dimensional space. Then, what we observe is a projection of an object in a lower dimension. Such a projection may result in observations that cannot be explained. For illustration purposes consider the trajectory on the top of Fig. 8. This trajectory is embedded in a three-dimensional space. If we project this trajectory onto a two-dimensional space we will obtain the result shown on the bottom of Fig. 8. Now the trajectory appears to intersect itself at one point. Every point in the two-dimensional picture corresponds to a point in the three-dimensional picture except for the intersection point, which represents two points in the three-dimensional picture. Thus, in a lower dimension those two separate states appear superposed. If something like this applies to our Universe, then we simply do not have the complete picture right. This in turn implies that we either do not know all the rules or we do not have enough information about the initial state. These possibilities create again randomness of the first and second kind. Finally, let us ponder on the many-words interpretation of quantum mechanics. According to this interpretation, the random quantum
processes cause the universe to split to as many copies as the possible outcomes. Thus, when our soccer player is ready to pass the ball, quantum effects in his brain lead to a superposition of many possibilities. All possibilities happen, but they happen at different universes. In our Universe what the player chooses (free will) appears as a slight randomness. The splitting of the universes can then be seen as a computationally irreducible process producing randomness from free will.

On now to the second law. Recall that the second law dictates that for irreversible processes the entropy increases. As we discussed earlier, this leads to the inevitable introduction of probability and randomness on the macro scale. Imagine you have a cup of warm water and a cup of cold water. To start with, you have an amount of information that specifies the difference in temperature between the two cups. You then pour the water from both cups into a pan. What do you get? Simply, you get lukewarm water. This is an irreversible process in which the entropy increases. Does this lukewarm water retain any of the original information? Apparently, not. You cannot say anything about the original temperatures any more. This demonstrates that irreversible process and entropy increase are associated with loss of information. The same phenomenon occurs in chaotic systems. When we measure an initial condition we may measure it with some uncertainty. Nevertheless, we do have some information about the initial condition. For example, we may measure the outside temperature with some error but we will have a pretty good idea of how warm or cold it is. Whatever information we have in the measurement of the initial condition is, however, lost in the future through the amplification of the uncertainty we have in that measurement. This is the same as saying that the system loses whatever information was supplied to it. It will then appear that the second law may be connected with randomness of the first and of the second kind. Alternatively, assuming that the system undergoing an irreversible transformation is an assembly of many, many individual particles, each one obeying some simple rules and interacting with every other particle, leads to stochasticity, the third kind of randomness.

Compelling arguments can therefore be made that the same ways for generating randomness suggested by pure mathematical systems may apply to real physical systems.

As we mentioned before, a key ingredient in all mechanisms of randomness is infinity. In that regard randomness may be thought as an infinite-dimensional (unrestricted) system, whereas rules can be thought as representing relatively low-dimensional (restricted) systems. Thus, while rules confine the dynamics, randomness acts without limits. This interplay creates most of the time ‘something’ in between. This ‘something’ in between is often referred to as complexity and it is the major characteristic of our Universe.

**Part 4: The role of randomness**

Imagine a group of small children in a playground. Small children have no fear of having an accident. They have not yet developed this feeling. As such they tend to move around irregularly changing activities and bumping into each other constantly. Such a set-up is always vulnerable to accidents. The question is, how do we minimize the chance for an accident? One way to eliminate accidents is to have one caretaker per child. In this case, the caretaker is on constant alert guiding each child in all its activities and making sure that no harm will come to it. This scenario amounts to complete determinism with no randomness allowed in the system. This solves the problem, but then all of us would have to become professional babysitters. Simply, this solution is not efficient. It requires too much effort. A more efficient solution would be to limit the area of activity (say, by putting a fence around the playground) and have a few caretakers supervise the children. In this case accidents may happen (and they do) but they will be much less frequent compared to the number of accidents when there are no rules or limits (randomness only).

Now imagine a parcel of air near the surface that is a perfect cube. As this cube begins to rise it expands, its relative humidity increases, and eventually becomes saturated. After that, as the parcel continues to rise, a cloud begins to form. But, what happened to our initial parcel during this process? As we all know the shape of a cloud is complex and no two clouds are alike. Given the fact that cubic clouds have never been observed, the initial cube simply gives way to some irregular structure. But why? One could imagine a process whereby each molecule in the original cube moves in such a way as to always form a cube. We could actually devise artificial rules that will have every molecule follow such an evolution. But can you imagine the effort that our atmosphere will have to make in order to achieve this? Nature will have nothing to do with processes like that. Instead, like the example with the children, the rules are set and within these rules the molecules are left to move randomly thus generating irregular cloud shapes.

Let us now consider different examples. Languages were one of the first necessities for humans. We simply had to communicate. But how do you think that
languages evolved? Take a little baby that begins to learn how to speak. What are the first words? If there is something universal in our cultures it is that babies all over the world begin muttering simple repetitive syllables like 'ba-ba,' 'ma-ma,' or something similar. They then proceed by becoming more elaborate. There is evidence that human language evolved similarly. In very primitive language words were made of very repetitive basic sounds. In fact, successful decipherment of ancient languages was based on finding repetitive units (Robinson 2002). This is true for music as well. The organization of music normally involves basic material that may repeat exactly or with variations, may alternate with other material, or may proceed continually to present new material. Composers strike a balance between unity and variety and all pieces contain a certain amount of repetition. As with languages, music may also have evolved from very simple repetitive musical blocks. In fact, early music was very repetitive.

The same has happened to the blueprint of life. Gene evolution is one of the most important aspects of evolutionary and molecular biology. Early in the 1970s the Japanese-American biologist Susumu Ohno advanced his ideas about a possible mechanism of evolution by gene duplication (Ohno 1970). In short, Ohno suggested that modern sequences arose from small pieces of genetic material (often called primordial blocks), which found a way to duplicate. Once this was possible, further duplication generated longer and longer sequences that led to the construction of genes. The same mechanism resulted in the generation of novel genes, by gene duplication, and new species by whole genome duplication. However, duplication alone does not produce any novelty. Because of that, another mechanism was at work together with the simple duplication. This mechanism is random mutation. As the primordial blocks duplicated they also mutated, meaning that they made slight changes in the repetition pattern. These random mutations are the key ingredient of one of the most powerful theories in history: Darwin’s theory of evolution. In short, according to Darwin, life evolved by natural selection and random mutations. Natural selection is the idea that individual species possess some variation that gives them an advantage over other species when it comes to survival. For example, imagine that in an isolated island populated by different species of birds, an environmental fluctuation has caused plants that produce small seeds to die, but plants bearing large seeds to survive. Because of some random mutation in their past some of the birds have developed big beaks. These birds have an advantage in picking up the large seeds compared to those birds that did not have this variation and remained with small beaks. Thus, the birds with the big beaks survive and the birds with small beaks become extinct. Environmental fluctuations and random mutations determined which species lived and which species died. The actual mechanism for this randomness in mutations is still an open question. It is widely believed that mutations arise from pure environmental factors. In this case the randomness may be due to stochasticity. Recent analysis of DNA sequences has also suggested that within this stochasticity some evidence can be identified suggesting that a component of this randomness may be connected to chaotic processes (Tsonis et al. 2002). Whatever the source of randomness, however, the point is that life with its entire splendor is the result of a very simple rule (duplication) plus randomness.

The similarities in the properties of languages, music, and DNA may indicate that all of them have employed a similar construction process in their evolution: repetition and mutations (randomness). And apparently, this process does not apply only to languages, music, and DNA. Given the plethora of self-similar structures in nature (which are created by repeating a certain operation over and over again), it would appear that there is something fundamental in evolving by copying or repeating an operation and modifying it at random, and that this process was adopted by both nature and humans in the early stages of evolution.

And why, you may ask, did such a procedure become the favored one? The study of languages, music, and DNA provides an interesting insight to this question. All three of them share a common property. They all transmit information. Furthermore, it is reasonable to assume that they all transmit information effectively and efficiently. Something that is effective and efficient uses the least amount of effort to do what it is supposed to do. There is no reason, for example, for nature to adopt a very complex and expensive mechanism to transmit information or to perform an operation. A simple and economical procedure would be much more desirable. And what can be simpler than repetition? It may not, thus, be surprising that once the art of repetition or copying was ‘learned’ it become a fundamental mechanism in nature and in human dynamics. But, since pure repetition does not create innovations, randomness was introduced to ‘spice’ things up. I do not mean to imply that this is the only modus operandi in the Universe or that other more complicated rules were not introduced later, but clearly simple rules and randomness do not just coexist but they synchronize to produce an efficient and economic Universe. It is interesting to note here that many mathematical systems obeying simple rules (such as cellular automata) have been reported, which copy or replicate themselves and which, through
replication, construct more complex structures (Langton 1986). The period doubling observed in the dynamics of the logistic equation is also an example of how duplication leads to complex behavior.

The above can be summed up by what Zipf called the principle of least effort or what I call the principle of minimum energy consumption (Zipf 1949). Mathematical and physical support for the minimum energy consumption of minimum dissipation principle is provided by the work of Ilya Prigogine. Ilya Prigogine was born in Moscow a few months before the revolution. His family left Russia in 1921 and, after spending a few years in Germany, settled for good in Belgium. His work in non-equilibrium thermodynamics (Prigogine 1980) won him the Nobel Prize in chemistry in 1977. Part of this work is the famous theorem of minimum entropy production, which states that when a system cannot reach equilibrium, but operates near equilibrium, the system settles to a state of minimum dissipation. Natural systems (and for that matter social economic and other systems) can operate at equilibrium, near equilibrium and far from equilibrium. From all these states we can argue reasonably that the most preferred state is the near-equilibrium state (far from equilibrium represents extreme situations and complete equilibrium means no more ability for changes). Accordingly, while the minimum energy consumption or minimum dissipation principle is not a universal principle it does apply to most phenomena observed in nature.

As Howard L. Resnikoff puts it in his book The Fusion of Reality, “Fermat’s classical variational principle of ‘least time’ and Maupertuis’ and Hamilton’s principle of ‘least action’ express the parsimony of nature in a mathematical form: the evolution of a physical system follows that path amongst all conceivable alternatives that extremizes, i.e. maximizes or minimizes, a suitable cost function such as time, action, or energy. Thus, the path of a ray of light through an optically inhomogeneous medium minimizes the time required to pass from the initial position to its emergent point.” (Resnikoff 1989). In the same issue, he argues that since the final state of an irreversible process is rather unpredictable (due to so many numbers of possible configurations), the final state of maximum entropy is a priori quite unknown. In this case, any measurement of that state yields the maximum information possible about the system (simply because before that there is no available information). In a sense, this represents a minimum effort to know something about the system. Therefore, maximization of entropy (and thus the second law) is consistent with the principle of least effort.

Summary

We started our adventure into randomness by looking exclusively at our formal mathematical system and we saw that even in this pure and strictly logical system one cannot do away with randomness. Rules and randomness are blended together and are engulfed in the notion of infinity. Staying within the mathematical system and employing simple mathematical models, we then discussed the three possible sources of randomness: randomness due to inability to find the rules, randomness due to inability to have infinite power (chaos), and randomness due to stochastic processes. Subsequently we expanded from the mathematical system to our physical world and we found out that randomness, through the quantum mechanical character of small scales, through chaos, and because of the second law of thermodynamics, is an intrinsic property of nature as well. We subsequently argued that the randomness in the physical world is consistent with the three sources of randomness suggested from the study of simple mathematical systems. Finally, we suggested the principle of least effort or the principle of minimum energy consumption as the underlying principle behind this combination.

We can thus conclude that no matter how randomness comes about, randomness and rules are bound together. They operate together. They synchronize together. They shape our Universe and produce the reality we see and feel every day in our lives. Randomness emerges as a property of the Universe. This synergy between rules and randomness makes them both equally important in the Universe. One cannot exist without the other. While rules impose boundaries, randomness acts between boundaries. They interweave together like facts and fiction in a historical novel. And overlooking this weaving is infinity, the one ingredient behind all mechanisms generating randomness. Possibly, it is what may make them one and the same thing.

I started with a quote by Aristotle, which I found appropriate to introduce the discussion. I would like to end it with another quote by Aristotle, which I find appropriate to the paper’s summary.

“Since nothing accidental is prior to the essential neither are accidental causes prior. If, then, luck or spontaneity is a cause of the material universe, reason and nature are causes before it.”

Aristotle, Metaphysics, Book XI, 8

Disclosure statement

No potential conflict of interest was reported by the author.
References